DOCUMENT RESUME

ED 454 857 IR 058 148

AUTHOR	Thomas, Sarah E.
TITLE	The Catalog as Portal to the Internet.
PUB DATE	2000-11-00
NOTE	18p.; In: Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (Washington, DC, November 15-17, 2000); see IR 058 144.
AVAILABLE FROM	For full text: http://lcweb.loc.gov/catdir/bibcontrol/thomas_paper.html.
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Access to Information; Cataloging; Electronic Libraries; *Gateway Systems; *Internet; *Library Catalogs; Library Role; *Library Services
IDENTIFIERS	Electronic Resources; OCLC

ABSTRACT

This paper examines the potential of the library catalog to serve as a portal to the Internet. The first section provides an overview of the development of the catalog, including the emergence of the union catalog, standardization of cataloging practice, MARC format, and the insufficiency of resources to catalog all the titles acquired by libraries. The second section addresses catalogs in the new millennium, including the variety of formats cataloged, enhancements in online catalogs to improve the quality of access, gateways to networked resources, database aggregations, OCLC's CORC (Cooperative Online Resource Catalog) service, and creation of a digital library architecture that embraces different formats and permits crossfile searching. The third section covers portals and catalogs, including definition of an Internet portal, differences between portals and catalogs, and deficiencies and benefits of portals. The fourth section discusses catalogs as portals, including goals and arguments in favor of libraries providing access to Internet resources. The fifth section offers recommendations for the future. (Contains 15 references.) (MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

The Catalog as Portal to the Internet

ED 454 857

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Sarah E. Thomas

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
B. Wiggins
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Final version December 2000

INTRODUCTION

"I don't do libraries," stated an engineering student last year at an Ivy League university, pleading with his professor to absolve him from an assignment requiring him to seek information in the campus library, presumably necessitating use of the library catalog. Increasingly, even at leading institutions of higher education, one encounters not just students, but also faculty and deans who assert that they get all the information they need through the Internet. In an interview with **D-Lib Magazine** editor-in-chief and digital library scientist Bill Arms reported in the **Chronicle of Higher Education,** Florence Olsen asks Arms: (1)

> Q. Do you think, within this decade, that digital libraries will replace traditional research libraries in most disciplines?
> A. I think it may be possible to have substantial research programs without access to conventional libraries.",

Arms then provides anecdotal evidence of a colleague who meets 80% of his information needs through open source documents. Another story in The New York Times was headlined "Choosing Quick Hits Over the Card Catalog," and reported: "Even though libraries are organized and easily navigated, students prefer diving into the chaotic whirl of the Web to find information."(2)

Libraries are awash in contradictions. Gate counts are up; circulation is down. While one set of constituents eschews traditional library services, another group pushes statistics for catalog searching steadily upward. Inside the profession, librarians engage in spirited debates about their role. In the face of doubters, librarians argue that only ignorant or naïve individuals would believe that the Web could satisfy all their information needs, particularly in the scholarly community. At the same time, they energetically acquire or license digital resources.

IR058148

With the addition of digital materials to the library's portfolio a debate about the role of the catalog has also developed. Should the catalog encompass all items that are considered part of a library's collection, even if those items are not physically held by the library? Should it even serve as a general gateway to

BEST COPY AVAILABLE

the entire Web? Proponents of the catalog and of libraries believe strongly that the catalog has enduring value and that it can evolve to be a useful tool for Web access, whereas critics do not foresee any role for the library catalog as a research tool for networked information.

This paper examines the potential of the catalog to serve as a portal to the Internet. It commences with a brief overview of the development of the catalog, details the attributes and limitations of library catalogs, and defines the concept of the portal. Finally, it offers proposals to respond to the dilemma of librarians about providing access to the expanding universe of information and knowledge.

LIBRARY CATALOGS AS A PORTAL TO KNOWLEDGE

It is always humbling to learn that something you regard as a great and very contemporary problem echoes an experience from the past. Recently a small tract documenting an address to the New York State Library School in 1915 by William Warner Bishop found its way to my desk. At the time of the address, entitled **Cataloging as an Asset,** Bishop was the Superintendent of the Reading Room in the Library of Congress. Bishop's observations merit reading, even after 85 years. He notes, "the library world has seen its shifting fashions, not to say its fads of the hour. And...the striking novelties are sure to attract a good deal of attention and to get themselves much advertised."(3) Relating the change in cataloging that occurred with the Library of Congress's successful implementation of the card distribution process, he suggests that this advance had lessened the perception of the importance of cataloging, and he declares: "Catalogs and catalogers are not in the forefront of library thought. In fact, a certain impatience with them and their wares is to be detected in many quarters. Shallow folk are inclined to belittle the whole cataloging business."(4) "I think I am safe in saying," he adds, "that most students in library schools would rather do anything else than take up cataloging on graduation."(5) Bishop goes on to deplore the catalogs of booksellers, created by non-experts, and he cites approvingly the value of the permanent contributions of catalogers in the enduring description of books. In his concluding remarks he is prophetic:

> We have just begun in America, an era of huge libraries, The average size is increasing very fast. Our large libraries are getting very large. They are being run for wide constituencies on broad lines. More and more the practical American spirit is seeking for coordination and cooperation. It is by no means certain that the card form of catalog will continue indefinitely as the chief tool of library workers. It is highly probable that selected catalogs will take the place of huge general repertories. Dimly one can see the possibilities of mechanical changes and alterations, of the use of photography, instead of printer's ink, possibilities of compression or even total change of form. Certainly our present card catalogs will require intelligent direction of the highest order to make them respond to the demands of readers, to the needs of the community. Changes such as these will require an intelligent and sympathetic oversight to insure their success. The librarians who will carry them out, who will guide and mold the development of cataloging, must perforce have been experienced and trained catalogers.(6)

When Bishop wrote, almost a century ago, the catalog was undergoing a transformation, and the cataloger was under siege. Cutter's **Rules for A Dictionary Catalog** had entered the librarian's canon, but Cutter's assumption was that the catalog referenced works held by a particular institution. While his goals for the catalog - being able to find all the works by an author, to find any work by title, to find all the editions of a work, and to find all works on a given subject, with the assumption being that the catalog referenced works held by a particular institution. Union catalogs expanded the function of the catalog to serve as an index to the holdings of multiple institutions, increasing their importance in the process.

Concomitant with the emergence of the union catalog was an increase in the standardization of cataloging practice. Early in this century, The Library of Congress revolutionized catalogs through the provision of printed cards. Over 675,000 titles were available by 1915 when Bishop wrote. Consider that in 1894, William Lane, Librarian of the Boston Athenaeum, conducted a survey of university librarians on cataloging practices as part of his preparation for writing a manual on library economy. Lane stressed in his cover letter: "Please indicate what different method (if any) from that which you actually follow you would prefer if you were settling the details of your catalogue afresh unhampered by past traditions." Survey question number 5 reads: "Do you follow pretty closely any code of catalogue rules? a. The A.L.A. rules. b. Cutter's rules. c. Linderfelts translation of Dziatzko. d. Columbia College Library or Dewey' rules. e. Jewett's rules. f. British Museum. g. Bodleian Library." Although a diversity of practice still abounds in 2000, the 20th century has seen major advances in the acceptance and employment of a number of cataloging and classification tools, including the Anglo-American Cataloguing Rules, the Library of Congress Subject Headings, the Decimal Classification system, and the Library of Congress classification system.

A key catalyst for the development of more uniform cataloging was the MARC format, created in the 1960s through major leadership and innovation at the Library of Congress. MARC enabled electronic dissemination of bibliographic records and engendered networks of libraries in such entities as OCLC and the Research Libraries Group. While initially MARC's power was felt in the economies realized through copy cataloging, first of records emanating from the Library of Congress, and subsequently, from original cataloging contributed through thousands of libraries, large and small, in the last two decades, MARC's potency has increasingly derived from unleashing the potential of the large-scale union catalog for resource sharing. It is a sign of our turbulent times that during a year in which the OCLC WorldCat database grew to 41,000,000 records, with 2.2 million bibliographic records added in fiscal year 1999, a session entitled "Is MARC Dead?" held in July at the American Library Association's annual meeting attracted an overflow crowd.

Standardized bibliographic records conveyed using the MARC format also led to the rise of local systems for the management of local library holdings. The OPAC (Online Public Access Catalog) assumed rising importance, and some librarians noted with dismay that the ease and convenience of the OPAC sometimes (often) lured searchers and lulled them into a complacency with results that were incomplete. Many institutions accelerated retrospective conversion of the card catalog to ensure that historical collections and fundamental publications acquired and cataloged prior to going online did not suffer from benign neglect. Some unconventional thinkers loaded records for titles not held by their

library, such as the catalog of the Center for Research Libraries, or UMI's Dissertation Abstracts, so that their clients might encounter resources, while not directly owned by their host organization, were readily accessible to them. RLG's Eureka databases and WorldCat were also considered logical extensions of the bibliographic universe available to students and researchers using a campus library.

A constant lament throughout the decades has been the insufficiency of resources to catalog all the titles acquired by libraries. Annual reports of librarians over two centuries are studded with references to accumulating backlogs. Open an annual report from any random year, turn to the section on cataloging, and almost certainly you will find a statement such as this one, drawn from the annual report of the Cornell University Libraries, 1946/47: "It is apparent from this listing of work to be done that the staff of the Catalog Department will have to be built up steadily to the point where it will be large enough to do the task assigned it. There is no other way in which the goal can be achieved. The backlog of work is very great and it will require a considerably expanded staff for a number of years to clear it up."(7) Administrators exhorted catalogers to be more productive, and in an effort to address the inexorable growth in workload as the volume of publications and acquisitions increased, catalogers, often led by the Library of Congress, introduced a number of collaborative programs to share cataloging and achieve economies. Their success in achieving enhanced productivity, though a combination of cooperative cataloging and enhanced tools, such as the cataloger's workstation, can be measured by noting that the number of catalogers employed in ARL university libraries has declined by 25% from 1990 through 1998 while the number of titles cataloged continues to rise.(8) Although some catalogers feared loss of job security if they successfully eliminated arrearages, new categories of materials to include in the catalog emerged to absorb any slack. Manuscript finding aids, guides to images, records for electronic resources, tables of contents, and other "non-book" materials competed for the attention of technical services specialists.

CATALOGS IN THE NEW MILLENNIUM

As we approach 2001, the information landscape appears to be considerably more complex than the one our predecessors populated. There is more information, the pace of change is more rapid, and the means and formats for communication are more diverse. What contribution does the catalog make in our quest to discover and retrieve knowledge? The catalog, at the level of the local institution, provides the information-seeker with bibliographic description and access to content imbued with several critical features. In addition to embodying Cutter's principles, the catalog has come to represent access to a collection deliberately shaped with a specific community in mind. This collection, by virtue of having been selected by bibliographers or some other structured process, is deemed to be of high quality. There is an implicit assumption that the works cited in the catalog are readily available for consultation. Furthermore because libraries have generally had a commitment to preserve and maintain those items they acquired, readers anticipate that a source identified today will be available in the future as well. Because they have been assembled according to standard practices and rules, by human intelligence, there is a high consistency in description, which in turn creates a high degree of predictability in results. This dependability generates an aura of trust. The user familiar with a catalog will have a high degree of confidence in the credibility of the sources contained in it. Another function of the catalog has been to link disparate materials. Until recently, the subject linkage has been chiefly among books, but in the past

few years, catalogs have begun to incorporate a variety of formats, including manuscripts, visual images, audio recordings, and now, in great numbers, digital objects. Finally, although catalog searching is a seemingly free good, with host institutions assuming the cost of maintaining local catalogs and paying for the subscription costs (but not free in the case of virtual union catalogs such as RLIN or OCLC.) Even the titles and proprietary information referenced by the catalog are more often than not purchased or licensed by a library and made freely available to its users. Recent enhancements in online catalogs have improved the quality of access. Some of the features found in state-of-the art catalogs are Web access, relevance ranking, more refined keyword searching, ability to limit by date or other information, and reference linking. Thus, the functionality of the online catalog is increasing, and its proponents are convinced that it can continue to remain an essential tool for the identification and location of documents and materials of importance for researchers. Today's OPAC holds records for books and journals, films, finding aids, audio recordings, computer files, maps, and graphic images, although the preponderance of surrogates are still for monographs and printed materials. As libraries subscribe to more and more online journals, full text documents, and other digital materials, catalog records refer to publications accessible to a community through a variety of authorizations. No longer are all the citations in a catalog to holdings owned by a library; pointing to materials served remotely has become commonplace. The purity of the principle that the local catalog provides access to materials held by the host institution has become diluted slightly to accommodate items selected for community use and readily accessible, although not physically controlled by the library. On the other hand, some librarians have balked at the introduction of certain types of electronic resources into the catalog, particularly those likely to have transient URLs or which require heavy maintenance. The catalog represents stability, dependability, reliability, and quality. Its holdings have not typically been ephemeral in nature. It goes against the grain for librarians to invest in the creation of an expensive and detailed bibliographic record if the resource for which it is a surrogate, is not likely to endure for the foreseeable future, if not permanently.

Recognizing that some patrons may prefer to connect directly with online resources without being routed through the catalog, some libraries have developed separate gateways to networked resources. These gateways facilitate access to electronic materials selected by the library by providing a single point of entry, by organizing them into categories, and using metadata, often derived from their catalog records, to assist users in locating networked resources. The gateway concept appeals strongly to those for whom speedy access to online resources is a priority, and it offers many of the desirable features of the catalog, since the bibliographic control over its contents is carefully managed by librarians. Although patrons have enthusiastically adopted the gateway at many organizations, there are some flaws in its design. Of concern for the library is the expense of maintaining synchronicity between the catalog and the gateway. Although clever programs enable the cloning of bibliographic records, entries in the catalog and the Gateway are not always identical. For example, Gateway records at Cornell are organized by simple subject categories, not by LCSH, and they contain less information than the AACR2 full MARC record in the catalog from which the Gateway entry is derived.

Another issue that has burdened catalogers has been the matter of database aggregations. The phenomenon of bundling journals or databases or other electronic materials into a single resource (JSTOR, ScienceDirect), has led to a heavy workload in those institutions which have chosen to analyze each individual title in an aggregation. The dynamic nature of these aggregations, in which titles are

added and dropped by the host provider on a continual basis, sometimes without notification, has significantly increased the labor entailed in adding, dropping, or modifying bibliographic records. Confoundingly, only a few suppliers of aggregations have to date seen the desirability of providing bibliographic records as a service, forcing each subscriber to repeat the effort of incorporating references to the titles they provide separately in their catalogs and/or gateways. This inefficient and wasteful situation has led to a variety of ameliorating initiatives.(9)

The Program for Cooperative Cataloging has worked with some vendors, such as EBSCO, ProQuest, CIS, and Gale, to stimulate the provision of wholesale bibliographic records to accompany subscriptions to its database aggregation.(10) These records can be loaded into a library's local system, increasing the standardization of access and saving local catalogers from the task of creating them from scratch or searching, downloading, and modifying for local use records existing in a national database This approach has had some success, but many publishers and vendors have lacked the staff expertise to create records of the quality expected by libraries. In some cases librarians have been unable to convince them that this is a service that would be worth the expense and effort of improvement.

In July 2000 OCLC put into production a service called CORC, the Cooperative Online Resource Catalog. Over 400 libraries are participating in the development of a Web-based product that uses a combination of automated tools and library collaborators to create a database of records to Web resources. Additionally, CORC includes an authority database, a pathfinder database, and a Dewey Decimal Classification Database. Users contribute URLs to the CORC database, and using automated tools, rapidly generate resource records. The system automatically suggests Dewey Decimal Classification numbers, keywords, and conducts authority checks, resulting in automatic authority control. URL maintenance is improved over its present, labor-intensive mode in local catalogs through the application of automated functions in concert with shared effort through the partners to distribute the workload. A library may export CORC records to a local catalog or gateway in either MARC or Dublin Core formats. OCLC will include CORC records in its WorldCat database.

Still another variation on the desire to manage access to Internet resources through the catalog, thereby maintaining the elements of predictability, authority, and stability of the traditional catalog, is the creation of a digital library architecture that embraces different formats and permits crossfile searching of materials cataloged, indexed, or otherwise controlled through a number of metadata schemes. Endeavor's ENCompass, currently under development, expands the view of the OPAC to enable users to direct a single query to multiple databases constructed using different encoding languages. The product is an open framework that uses metadata standards such as Dublin Core, EAD (Encoded Archival Description), and TEI (Text Encoding Initiative) to provide access to full-text resources, finding aids, and other digital objects that the ENCompass host has identified as relevant to its user community. ExLibris is developing a similar product called MetaLib. VTLS has developed a three-part approach, "Library Automation in 3V," which includes a system to handle internal library processes, a second component to support digitization, indexing, linking, and access of multimedia materials, and a third part to facilitate integration with external sources and technologies. These initiatives offer promise for the immediate future for effective access to a broader range of materials.

As noted, libraries have struggled for years to stay ahead of the rising tide of printed publications as they labored to provide bibliographic control. The Library of Congress, for example, heroically reduced its backlog of monographs over the past decade. Yet, despite some measure of success through a combination of cooperative initiatives, new technological advances, and occasional staff increases, the essential problem of cataloging or otherwise describing and analyzing the world of knowledge has remained an enormous challenge. As print indexes morphed into online databases, some voices admonished that libraries ought never to have allowed the indexing business to migrate from their domain into the commercial sector in the 1930's, since we now see the price we have to pay for access to these valuable resources escalate. The penetration of visual culture into scholarly activity necessitates improved access and more widespread dissemination of records about visual images. Other formats and materials, such as manuscripts and audio transcriptions, have ascended in importance. The interest in these materials, which have often been sequestered in special collections, has risen in part as digital technology has facilitated their visibility and accessibility. Although the backlogs in these formats (manuscripts, music, photographs, moving images, sound recordings, and maps) were even more egregious than those of books and serials, LC has sought to increase formal control over them in the past few years, and other institutions have raised the priority of their special collections as well. The numbers remain daunting, however. At one large research library, the task of converting all existing finding aids using EAD and gaining descriptive control over its entire collection of manuscripts was estimated to exceed $3 million, and since its technical services operations, using its present methodology to organize its collections, is chronically understaffed, it expected to increase this figure by a quarter of million dollars per year, taking into account the rate of new acquisitions.

During the same period that libraries have been asserting control over their backlogs of printed publications and have been shining their light on the hidden resources found in archives and special collections, the World Wide Web sprang to life. Few people had the clairvoyance to anticipate its astonishing growth and vitality. Today it registers 1.5 million new pages per day, and with a present size estimated to be in excess of 2 billion pages, it represents a major challenge to the traditional library practices. As there is mounting evidence that students, faculty, researchers, and the general public are making the Internet their information resource of the first and last resort, library values of careful selection, standardized description, and enduring access to publications are questioned as both costly and futile. A common assertion by those conversant with the Web is that library tools such as AACR and MARC won't scale in the Web environment. One digital library specialist has advanced the theory that an Internet search engine, such as Google, could replace the expensive, labor-intensive aspects of librarianship, obviating the need for catalogers, reference librarians, or selectors, or at least significantly reducing the university's dependence on them. As Bill Arms ventures in an article entitled Automated Digital Libraries:

> Quality of service in automated digital libraries will not come from replicating the procedures of classical librarianship. More likely, automated libraries will provide users with equivalent services that are fundamentally different in the way they are delivered. For example, within the foreseeable future, computer programs are unlikely to be much good at applying the Anglo-American Cataloguing Rules to monographs. But cataloguing rules are a means to an end, not the end itself. They exist to provide services to users, notably

> information discovery. Automatic methods for information discovery may not need traditional cataloging. The criterion for evaluating the new methods is whether the users find what the information they require.(11)

PORTALS AND CATALOGS

With the Web estimated to be increasing by 10 million pages weekly, the task of indexing Internet resources is clearly gargantuan, and not something that can be accomplished by even the most industrious honeybee hive of catalogers. Instead of relying on the catalog to identify and retrieve relevant web pages, users have turned instead to Web portals. The term "portal" has gained currency recently as an entry point to the web. Traffick, the Guide to Portals, www.traffick.com traces the portal's antecedent to the search engine or directory service that began to take advantage of the millions of site visits they received daily. The search engine sites recognized commercial potential by adding features that would entice repeat visits and encourage the pursuit of particular links that would advantage their partners or advertisers. In a Princeton resource published by the InSide Gartner Group, Debra Rundle offers this definition of an Internet portal:

> Internet portals originated as the librarians of the Web. The word "portal," meaning "door," has been used to characterize Web sites commonly known for offering search and navigation tools. Circa 1996, a portal was used to catalog the available content from the Internet, acting as a "hub" from which users could locate and link to desired content. Their business models consisted solely of selling advertising banner space and directing Web surfers to their desired destinations successfully (to ensure repeat business).
>
> Now portals are more than just a launching pad to content at other sites. They offer a broad array of online resources and services. Although there is no single model for what constitutes a portal, all portals offer at least five core features: Web searching, news, reference tools, access to online shopping venues and some communication capabilities (i.e., free E-mail and chat)(12)

Howard Strauss, Manager of Academic Applications at Princeton, defines a portal as a "gateway to web access" or "a hub from which users can locate all the web content they commonly need." He asserts that mandatory features of a portal include personalization, search, channels, and links, and that desirable elements are customization, role-based models, and workflow.(13)

According to Looney and Lyman, "portals gather a variety of useful information resources into a single, 'one-stop' Web page, helping the user to avoid being overwhelmed by 'infoglut' or feeling lost on the Web."(14) They estimate that 89% of the approximately 58 million Web users in the U.S. frequent portals, and they subdivide portals into categories such as the consumer portal (directory sites such as AOL, Yahoo!), community portals, which collect and organize information relating to a particular subject or interest group, vertical portals, which are often a unified site created by a particular service provider and organized on a special business topic (ETRADE), and an enterprise portal which provides a channel for intranet and external data for a corporation or university.

Portals differ significantly from library catalogs in several key ways. Like the catalog, they are built around the concept of a community, although a considerably larger body of users than the typical library catalog user. Unlike the catalog, they integrate all manner of information in their scope, rather than concentrating exclusively on "published" information. Frequently they contain a strong commercial element, with advertising prominent on their pages, and often affecting the display of search results. The search engines they employ use programs to harvest URLs and generate responses. Search queries yield large response sets, often in the thousands, and the items retrieved include duplicates, false drops, results skewed by deliberate manipulation of terms by their authors, materials of dubious heritage: in short a vast flea market of junk, collectibles, and genuine antiques. Large numbers of the URLs retrieved lead to dead ends, where the site has moved or dropped off the face of the earth or where the information has ceased to be updated. Users spend an inordinate amount of time sifting through the vast finds, often failing to locate the best resource.

The Internet portals are rife with deficiencies. They lack the very characteristics which are the virtues of the catalog. Their value, on the other hand, is lacking in the catalog. The information they access is prolific, and is often very current. With the hyperlinked aspect of the Web, it is easy to move from document to document, and the generous amount of full-text resources allows the user to mine very specific terms. There is vastly more audio and visual data available for consultation. The user can conduct her research without the inconvenience or disruption of leaving her computer, and she can readily cut and paste the results of her searches into her own documents. Result sets are ranked by relevance, and can be tailored to personal specifications. These characteristics, along with many other positive features of the Internet, excite an enthusiasm for the Internet that outweighs the deficiencies for large numbers of the population of information seekers.

Is it possible to merge the best of the portal with the strongest attributes of the library catalogs? In 1999 several library leaders began exploring a concept of a library portal in a series of structured discussions. Jerry Campbell, CIO and Dean of University Libraries, University of Southern California, a participant in these sessions, has described the proposal for a "scholars portal" in a white paper prepared for the ARL annual membership meeting in May 2000. (15) According to Campbell, the "scholars portal would promote the development of and provide access to the highest quality content on the web...." The scholars portal would foster standards and provide cross database searching. In addition, to the provision of quality content appropriate for scholarly discovery and research, it would offer affiliated services, such as reference services. The scholars portal would stand in clear opposition to the "information.coms" with their indiscriminate content and commercialized milieu.

CATALOGS AS PORTALS?

How could a library catalog serve as a portal to the Web? One thing that it could never do is function as the sole gateway to all Internet resources. Even a collaborative endeavor such as CORC could not fulfill this role, as the quantity and diversity of Web resources defy such comprehension. Even if one were to limit the candidates for control to the high quality resources contemplated as links in the "scholars

portal", one should assume that the catalog would serve as only one point of access to web resources by users, who would likely have several other portals they would consult, based on their affinity groups.

Instead of striving for comprehensiveness, the goal of the catalog as portal must be to increase the ability of a community of users to meet their information needs by doing as much "one-stop shopping" as possible. By including access to web resources in the catalog, libraries would be extending to some Internet materials the same level of control that they have traditionally provided for analog formats. They would convey, through their integration in the online catalog, the credibility conferred through an affirmative selection by an intelligent being. The presence of a citation in a catalog has come to signify for the user that the source discovered is readily obtainable, that it has been chosen for its relevance to past and present foci of the community of which the searcher is a member; that the material possesses authenticity, in that the rigor of the selection process vouches in some way for its scholarly value; and that the document consulted today will be persistently available for future examination. The wrapper of the catalog conveys respectability on its contents. Readers recognize that the texts and documents referenced in the catalog represent a diversity of viewpoints, but that the universe of publications on a particular topic has been screened (or some portion of that universe) to separate out those objects which have traditionally had the greatest value for a particular constituency. In the past, those publications have had a heavy concentration of highly edited, peer-reviewed, frequently cited publications, and the virtue of the catalog for discovering materials meeting these and other unwritten standard of quality still continues. The director of Xerox's Palo Alto Research Center, John Seely Brown, parsed the difference between the web and a library, stating:

> On the Web, most information does not have an institutional warranty behind it, which really means you have to exercise much more judgment. For example, if you want to borrow a piece of code or use a fact, you'll have to assess the believability of the information. If you find something in a library, you do not have to think very hard about its believability. If you find it on the Web, you have to think pretty hard.(16)

There is a strong argument for libraries providing access to (some) Internet resources for their clients. By creating a mechanism that offers a particular subset of information seekers the ability to search citations (and more) to a pool of information that includes all formats, libraries can offer a service that increases the productivity of the searchers. They can forge a link between past knowledge, as collected and curated in library and archival repositories, and emerging ideas, as manifested in a variety of media, in a way that a search engine which restricts itself to the URL's of web pages cannot. And libraries can permit and facilitate the discovery and use of proprietary information that is not open to the independent Web searcher using a commercial portal. This licensed content may not even be located through the search engine serving that portal because of the security wall the content provider has erected to defend its property.

RECOMMENDATIONS FOR THE FUTURE

Having justified the creation of a mechanism managed by libraries to support access to Internet

resources, the next question becomes: should the catalog serve as the portal to the Web? Are the tools used to build the catalog appropriate for description of Web resources? This conference will examine the flexibility of AACR2, other metadata schemes, MARC, and other standards that librarians have commonly employed to describe, categorize, and communicate information about materials held in libraries or identified by libraries as relevant to their users. These tools are good and durable instruments, and over the years I have resented comments such as "MARC costs too much to apply," or "AACR is too complicated." In themselves, these tools are not insurmountable hindrances, and in fact, they have much good to contribute to our ability to organize knowledge. Yet, at the same time, as a library administrator, I am apprehensive about applying the same standards and procedures we are using for books and journals to Internet resources.

As we move into the 21st century, we must consider reorienting ourselves and rethink the way in which we provide access to information and knowledge. Our familiar aids, such as AACR2, should be probed for the values and basic principles of organization they yield. The IFLA Functional Requirements for Bibliographic Records contribute substantially to our understanding. But we must conceive of new ways to accomplish our goals by building actively on the past while freely abandoning rules that restrain us and readily adapting new technologies. Michael Gorman has suggested a tiered approach to the description of publications that takes into account the quality of the material being described, with a progression from AACR2 through Dublin core to keyword search indices.(17) This is sensible counsel, and provides a path from the present to the future.

One of the biggest challenges facing us is the sheer volume of material that is worthy of scholars' consideration. David Levy has noted: "There is a growing awareness of attention as a highly limited resource, stemming in part from the realization that an abundance of information, good though it is in many ways, is also a tax on our attention."(18) The filtering and organizing done by libraries has the potential to serve as a labor-saving device and productivity tool for researchers in a way that is now, in the delight over the fertility of the Web for expression, only dimly appreciated by a few. But, like the enthusiasm for the automobile that propelled the acquisition of vehicles and the construction of highways but which has spawned today concern about sprawl and congestion, the Internet will seek regulation and traffic calming devices. The library catalog, or some permutation of it, can help.

To accomplish this, we must look at a number of possible changes in the way we do our business:

1. We should decisively reduce the amount of time we devote to the cataloging of books in order to reallocate the time of our bibliographic control experts to provide access to other resources, especially Internet resources, but also unique primary resources and other analog format materials.

2. In order to reduce the time spent cataloging books, we will need to investigate and implement a combination of the following :

   Using the PCC core bibliographic record (see www.lcweb.loc.gov/catdir/pcc/corebook.html)

Using Dublin core or a modification thereof

Accepting copy with little or no modification from other cataloging agencies, including vendors

Working with publishers, authors, and software developers to encode publications in a standard way that permits the generation of metadata from digital objects through the use of software programs

Increasing collaborative efforts nationally and globally so that publications are cataloged according to mutually acceptable standards in a timely fashion and once only.

3. To increase the functionality of the library portal/catalog, libraries need to:

Increase the scope and coverage of materials

Ensure timely access to publications

Increase the level of access from citation to full-text or increasing degrees of granularity.

Incorporate features such as reference linking, recommended titles (others who liked this title also liked:), relevance ranking, customization, and personalization that make portals so captivating

4. To ensure success, libraries shouldn't go it alone. Libraries should:

Collaborate with other libraries in a coordinated plan for the acquisition, creation of metadata, access, and preservation of materials available through portals.

Define a clear path from the local library portal to the larger scholars portal

Partner with developers of portals and search engines to share expertise in a constructive way, drawing on the best each has to contribute to the goal of effective access to information

5. Don't hide our light under a bushel. Libraries should:

Advertise the features of the discovery database, a hybrid combining some of the best features of the catalog and the portal, using local and global outlets.

Quantify the value of the laborsaving features of the portal/catalog for the community of potential consumers and for those administrating the organizations who subsidize them and

stand to benefit from them

Seek new revenue (from partner portals?) to be able to expand their scope and accomplishments

Conduct and publish research documenting improved results through use of the catalog (saves time, finds more appropriate materials; titles found are accessible, etc.)

We presently lack the resources to provide access to all the information we would like to include. In addition to changing our practices to be able to expand coverage with existing funding, we should seek additional support through Congress for LC's leadership and participation, from granting agencies such as NSF and NEH to support research and pilots in the development of metadata harvesting software, crosswalking and associated access capabilities. We should seek the support of the organizations such as OCLC, RLG, and the Digital Library Federation for research in improving means of access and in fostering collaborative programs. We should work within our geographic regions, our consortia such as CIC or NERL, and other networks to accelerate the acceptance of best practices and to create linked catalogs with reinforcing document delivery and coordinated archival responsibilities. We should work within our associations and our home institutions to build a public awareness of and appreciation for the service provided by the catalog and its creators. This contribution should be documented with both the tangible contribution to members of the host institution and the intangible value of the public good the catalog represents.

The catalog can serve as a portal to the internet if the catalog is reinterpreted to be an information service which registers in a systematic arrangement those publications and documents of interest to a particular community, regardless of the form in which they appear. This discovery and access tool may exploit a variety of metadata schemes to locate materials, but it imparts unity, predictability, authority, and credibility to search results through the efforts of expert knowledge managers and the application of principles, policies, and practices of their devising. In the short term, we can expand the catalog to be more inclusive and flexible. In the near future, however, we should expect a hybrid which will adopt some of the superior features of the catalog, but which will employ an increasingly sophisticated technological infrastructure to increase the yield for information seekers. This information management tool will have evolved from the catalog and will be influenced by what we today call the portal, but it will likely have a newly coined name to represent a new concept. This "Open Sesame" service will incorporate the trusted aspects of the catalog, granting the searcher access to a realm rich with quality resources which she can easily locate and which more often than not hits the target of her needs. At the same time, the lode will yield an array of up-to-date data covering a breadth of formats and a depth of detail.

To achieve this new information medium, we will have to have the courage to risk change and to explore unfamiliar territory. Ultimately, we should figure out a new construct in which we will devote a greater proportion of our resources to providing access to materials previously left uncataloged, but which today are an important aspect of the information landscape. Accomplishing this will require a fairly dramatic shift in attention in libraries. Reallocating 10% of our cataloging resources to address this future direction

may be insufficient, but even this small amount could make a noticeable difference in thinking about which attributes of the catalog have the highest priority to apply to the broader range of materials and to considering new ways of attaining the desired goals. It might be necessary to alter the way all items are processed to redirect 10% of our resources, or we might continue to treat a certain number of materials as we have, but drastically reduce the fullness of the record for others. One thing is certain: ten percent is only a beginning. We will have to organize ourselves quite differently to provide service that is meaningful, relevant, and useful for scholars and students, and if we do not do this quickly, even our worthwhile contributions will be overlooked by many whom we could aid.

The new model of information tool should draw on the wisdom of the librarian in organization, but will use the savvy of the programmer to produce the most cost-effective and accurate results possible. In its ideal realization, the successor to the library catalog will express its virtues, but will supplement them with many new features made possible through technology. The best way to accelerate the transformation of the catalog into this new entity will be to participate openly and substantively in the design of new systems into which we can transfer certain enduring values.

Notes

1. Florence Olsen, "Logging in with...William Arms: 'Open Access' is the Wave of the Information future, Scholar Says,"The Chronicle of Higher Education, Friday, August 18,2000."
2. Lori Leibovich, "Choosing Quick Hits Over the Card Catalog," The New York Times, August 10, 2000, G1, G6.
3. William Warner Bishop, Cataloging as an Asset, Baltimore: The Waverly Press,1916, p. 4.
4. Ibid., p. 7
5. Ibid., p. 18
6. Ibid., p21-22.
7. Cornell University Libraries, Annual Report 1946/47, p. 15
8. ARL:A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC, 208/209 Feb. Apr 2000, p.5
9. Karen Calhoun and Bill Kara, "Aggregation or Aggravation? Optimizing Access to Full-Text Journals, ALCTS Online Newsletter, (Spring 2000). www.ala.org/alcts_news/v11n1/index.html
10. PCC Standing Committee on Automation Task Group on Journals in Aggregator Databases, Final Report (January 2000), lcweb.loc.gov/catdir/pcc/aggfinal.html
11. (William Y. Arms, "Automated Digital libraries: How Effectively Can Computers Be Used for the Skilled Tasks of Professional Librarianship?" D-Lib Magazine, July/August 2000, www.dlib.org/dlib/july00/arms/07arms.html
12. www.princeton.edurundle/PrincetonPortal.htm Document #IGG-03241999-02, 24 March 1999).
13. www.cren.net/...techtalk/events/campusportals.html
14. Michael Looney and Peter Lyman, Portals in Higher education: What are they and What is their Potential, EDUCAUSE Review, July/August 2000, p.30.
15. Jerry D. Campbell, "The Case for Creating a Scholars Portal to the Web: a White Paper," prepared

for the Association of Research Libraries, April 13, 2000,www.arl.org/newsltr/211/portal.html
16. Lawrence M. Fisher, "An Interview with John Seely Brown, Strategy & Business, Issue 17, Fourth Quarter 1999, p. 93-94
17. Michael Gorman, "Metadata or Cataloging? A False Choice." Journal of Internet Cataloging, v.2, no. 1 1999, p. 5-22
18. David Levy, "I Read the News Today Oh Boy: Reading and Attention in Digital Libraries, Proceedings of the 2nd ACM international conference on digital libraries, July 23 - 26, 1997, Philadelphia, PA USA" p. 202-211 (p. 202)

Library of Congress
December 21, 2000
Comments: lcweb@loc.gov

Conference Home Page

What's new

Greetings from the Director for Cataloging

Topical discussion groups

NAS study and 2 articles from the LC staff Gazette

Conference program

Speakers, commentators, and papers

Conference sponsors

Conference discussion list

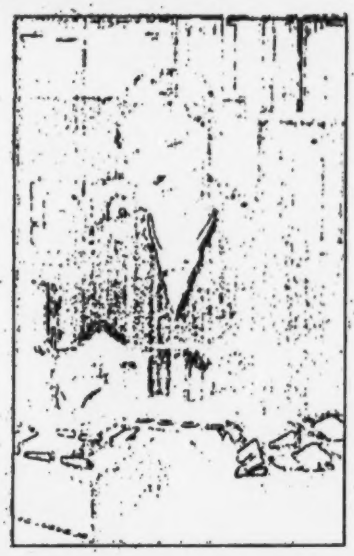
Logistical information for conference participants

Conference Organizing Team

Bicentennial Conference on Bibliographic Control for the New Millennium

Confronting the Challenges of Networked Resources and the Web

sponsored by the Library of Congress Cataloging Directorate



Sarah E. Thomas

University Librarian
201 Olin Library
Cornell University
Ithaca, NY 14853-5301

The Catalog as Portal to the Internet

About the presenter: Sarah Thomas came to Cornell University in August 1996 as the Carl A. Kroch University Librarian. In a career spanning over 25 years, Thomas has cataloged books in Harvard University's Widener Library, taught German at The Johns Hopkins University, managed library coordination at the Research Libraries Group (RLG) in California, held a Council on Library Resources Management Internship at the University of Georgia, served as the Associate Director for Technical Services at the National Agricultural Library, and directed both the Cataloging Directorate and the Public Service Collections Directorate at the Library of Congress. At Cornell, she provides leadership for the 19 libraries that make up the University's library system, managing a staff of over 500 employees and 600 students. The Cornell University Library holds over 6.7 million volumes and acquires and catalogs over 100,000 titles annually.

Thomas has had a long-standing interest in information technology. She currently serves on the Executive Steering Committee of the Digital Library Federation, and she frequently speaks or writes on the topic of digital libraries. In May 1998, she was appointed a member of the New York Regents Commission on the Future of Library Services. She is a life member of ALA and serves as the chair of the Access to Information Resources Committee of the Association of Research Libraries (ARL) as well as a member of the ARL Board. She is a member of the Board of RLG and serves on advisory councils to several university libraries, including Harvard, MIT, and Washington University. Thomas earned a Ph.D. in

Cataloging Directorate Home Page

Library of Congress Home Page

German literature from The Johns Hopkins University in 1983, writing her dissertation on the topic: "Hugo von Hofmannsthal and the Insel-Verlag: A Case Study of Author-Publisher Relations." She received her bachelor's degree from Smith College in 1970 and an MSLS from Simmons College in 1973.

Full text of paper is available

Summary: For well over a century, the catalog has served libraries and their users as a guide and index to publications collected by an institution. Charles Cutter's principles--to enable a person to find a book of which either the author, title, or subject is known; to identify all the titles held by the library on a given subject or genre, or written by a given author; or to assist in the choice of a book by edition or character--still motivate the practice of cataloging and continue to offer a framework for organization that is relevant in the world of the Internet.

The attributes of the catalog that have made it a valuable resource are desirable traits in any information management tool. Library catalogs provide those consulting them with a degree of predictability, authority, and trusted selectivity. The library catalog user has traditionally assumed that items listed in the catalog were carefully chosen to support an institutional mission and that they were available for her inspection. Internet portals, gateways to the Web, like the catalog, offer access to a wide range of resources, but differ from the catalog in a number of ways, perhaps most significantly in that they facilitate searching and retrieval from a vast, often uncoordinated array of sites, rather than the carefully delimited sphere of the library's collections. Web information has proven much more volatile, ephemeral, and heterogeneous.

Can we re-interpret the catalog so that it can serve effectively as a portal to the Internet? Is the catalog the appropriate model for discovery and retrieval of highly dynamic, rapidly multiplying, networked documents? Until relatively recently, the catalog has been the dominant index to published literature for library users. Web portals are rapidly usurping this primacy. Libraries today are struggling as they strain to incorporate a variety of resources in diverse formats in their catalogs and to maintain centrality and relevancy in the digital world. This paper will examine the features of the catalog and their portability to the Web, and will make recommendations about the Library catalog's role in providing access to Internet resources.